

Half-Spectral Space-Time Covariance Models

Michael T. Horrell & Michael L. Stein
University of Chicago

May 7, 2015

Abstract

We develop two new classes of space-time Gaussian process models by specifying covariance functions using what we call a half-spectral representation. The half-spectral representation of a covariance function, K , is a special case of standard spectral representations. In addition to the introduction of two new model classes, we also develop desirable theoretical properties of certain half-spectral forms. In particular, for a half-spectral model, K , we determine spatial and temporal mean-square differentiability properties of a Gaussian process governed by K , and we determine whether or not the spectral density of K meets a regularity condition motivated by a screening effect analysis. We fit models we develop in this paper to a wind power dataset, and we show our models fit these data better than other separable and non-separable space-time models.

Keywords: Space-Time Processes, Spectral Density, Fourier Transform, Covariance Function

1 Introduction

Continuous natural phenomena in space or space-time are often modeled statistically as Gaussian processes. Generally, a Gaussian process model requires specification of a mean structure and a covariance or dependence structure. Since covariance models must be positive definite, mathematically valid dependence structures can be quite difficult to define. Therefore, a primary focus in defining new Gaussian process models is on defining valid covariance functions. A covariance function, K , can be defined explicitly or implicitly in a number of ways. For example, for continuous $K(x)$, where $K(x) = \int \exp(i\xi^T x) F(d\xi)$ with $x, \xi \in \mathbb{R}^{d+1}$, specifying F or the spectral density, f (assuming it exists), where $F(d\xi) = f(\xi)d\xi$, will represent and characterize dependence structure in a Gaussian process. In this paper, we explore an alternative representation of K we call the half-spectrum. We show half-spectral models are easy to define naturally, and we show many half-spectral models have several nice modeling properties.

For a stationary, continuous space-time covariance model $K(s, t)$ with $s \in \mathbb{R}^d$ and $t \in \mathbb{R}$ with integrable spectral density, $g(\lambda, \omega)$, where $\lambda \in \mathbb{R}^d$ and $\omega \in \mathbb{R}$, we can represent K with

g via Fourier transform,

$$K(s, t) = \frac{1}{(2\pi)^{d+1}} \int_{\mathbb{R}} \int_{\mathbb{R}^d} g(\lambda, \omega) \exp(is^T \lambda + it\omega) d\lambda d\omega. \quad (1)$$

A half-spectral representation of K can be obtained by performing one of the two integration steps in (1) with respect to either λ or ω . After integrating over either λ or ω , K is represented by the Fourier transform of an expression relating distance and frequency pairs: (s, ω) or (λ, t) . We therefore call covariance representations of this type either half-spectral models in time (if in terms of (s, ω)) or space (if in terms of (λ, t)). In this paper, we study half-spectra in time of the form $f(\omega)\mathbb{C}(s\delta(\omega))\exp(i\theta(\omega)\phi^T s)$ defined such that

$$K(s, t) = \int_{\mathbb{R}} f(\omega)\mathbb{C}(s\delta(\omega))\exp(i\theta(\omega)\phi^T s)\exp(it\omega)d\omega. \quad (2)$$

Here f is the spectral density of $K(0, t)$, the temporal covariance function at a single spatial site; \mathbb{C} is any valid correlation function in \mathbb{R}^d with a spectral density; and δ is an even positive function that determines the space-time interaction properties of the model. The odd, real function, $\theta(\cdot)$, and the unit vector, $\phi \in \mathbb{R}^d$, determine space-time asymmetry in K by specifying a phase relationship. By Theorem 1 in Stein (2005b), we know K thus defined is a valid covariance model.

Space-time asymmetry can in many cases be thought of in terms of movement of a process through time. Gneiting (2002) first introduced the term fully-symmetric to describe covariance models that have the following property: $K(s, t) = K(-s, t) = K(s, -t) = K(-s, -t)$ for all $s \in \mathbb{R}^d$ and $t \in \mathbb{R}$. For models of physical processes, full-symmetry can be a strong assumption; thus, θ and ϕ are important model building tools. For much of this paper, we primarily consider the space-time symmetric case with $\theta(\omega) = 0$; hence, we mainly study half-spectra of the form $f(\omega)\mathbb{C}(s\delta(\omega))$. However, as we show in Section 2, some important modeling properties of K will hold, independent of θ or ϕ . Restriction of our analysis to models defined by $f(\omega)\mathbb{C}(s\delta(\omega))$ should therefore not be seen as a simplification.

Cressie & Huang (1999) first used half-spectral representations in space to help define several non-separable, closed-form space-time covariance functions. Gneiting (2002) used half-spectral models as a stepping stone to extend much of the work in Cressie & Huang (1999). Gneiting (2002) also focused on establishing a large and flexible class of closed-form covariance functions. Stein (2005b) showed half-spectra in time can be fit directly to certain data structures using a multivariate Whittle likelihood approach. Kent & Mosammam (Submitted) also focus on the class given in (2) and produce model fitting and inference strategies. Given the general usefulness of half-spectral models, it is natural to push their development. We go a step further than these authors by developing several properties of certain half-spectral forms.

The fully-symmetric, half-spectral form we study, $f(\omega)\mathbb{C}(s\delta(\omega))$, is quite flexible and lends itself to new interpretations. Purely temporal properties of these models are determined by f , and purely spatial properties are determined by the real integral $\int f(\omega)\mathbb{C}(s\delta(\omega))d\omega$. This real integral can also be seen as a mixture of covariance functions with varying ranges or simply as an expectation. These models also serve as one fairly direct way to extend common purely temporal models to the spatio-temporal domain. The space-time interaction

properties of these models are determined by δ . When δ is constant, temporal and spatial marginal covariance functions can be specified completely independently. In this setting, $K(s, t)$ is proportional to a product of its spatial and temporal marginal covariance functions: $K(s, t) \propto K(s, 0)K(0, t)$. A model with this property is called a separable model, and like Cressie & Huang (1999), Gneiting (2002) and several others, we avoid separable models because they have been criticized as unnatural (Kyriakidis & Journel 1999). Finally, if the spectrum of \mathbb{C} is known, the full spectral representations of $f(\omega)\mathbb{C}(s\delta(\omega))$ or $f(\omega)\mathbb{C}(s\delta(\omega))\exp(i\theta(\omega)\phi^T s)$ can be determined straightforwardly; thus, full spectral diagnostics and considerations can be easily applied to half-spectral models of this type.

We use the flexibilities of $f(\omega)\mathbb{C}(s\delta(\omega))$ and $f(\omega)\mathbb{C}(s\delta(\omega))\exp(i\theta(\omega)\phi^T s)$ to build new classes of models with several desirable modeling properties. Stein (1999) has established the tremendous importance of the degree of mean-square differentiability (or smoothness) of a process on predictions and assessing uncertainty in these predictions. We therefore work to develop models that have flexible degrees of smoothness in space and time. We also consider carefully the space-time interaction properties of the models we develop. Again, we avoid space-time separability, but we additionally attempt to define models that have “natural” space-time interaction properties. There is less research on properties of space-time interaction in space-time models; hence, “natural” space-time interaction is difficult to define without a specific application in mind. However, Stein (2011) provides one condition that we might want to assume in practice. Specifically, Stein (2011) argues for the following condition on the full spectrum, g :

$$\lim_{\|(\lambda, \omega)\| \rightarrow \infty} \sup_{\|(u, v)\| < R} \left| \frac{g(\lambda + v, \omega + u)}{g(\lambda, \omega)} - 1 \right| = 0 \quad (3)$$

for all finite R . The limit $\|(\lambda, \omega)\| \rightarrow \infty$ indicates any path to infinity may be taken. In particular, either λ or ω may be bounded or fixed as the other coordinate goes to infinity. Essentially, the condition in (3) requires g to flatten out in a relative sense at high values of λ and/or ω .

Stein (2011) justifies (3) by considering interpolation properties of models determined by g . Specifically, Stein (2011) considers asymptotic behavior of best linear unbiased interpolations using points near to and far from an interpolation site. Stein (2011) proves that if g satisfies (3) and has further mild properties, then as the nearby points approach the interpolation site, interpolation using only the information in the nearby points becomes optimum in the sense that adding the information from the far points does not improve predictions. This broad concept is called the screening effect (Chilès & Delfiner 1999). The theorems in Stein (2011) are not completely general, but they suggest that a space or space-time model with a spectral representation that does not follow the condition in (3) may have undesirable interpolation properties. We therefore use this condition to guide development of our classes of half-spectral models. Hence, the models we develop here may have more natural interpolation properties and, by extension, more natural space-time interaction.

As further confirmation that the condition in (3) may describe natural models, a number of physically derived Gaussian process models will satisfy (3). For example, models defined using stochastic partial differential equations will often meet (3) when white noise forcing terms lead to processes with integrable spectral densities. In particular, Gaussian process

solutions to stochastic versions of convection-diffusion equations will satisfy (3) (Lindgren et al. 2011).

Several classical “unnatural” models do not meet the condition in (3). Separable models in particular do not follow (3). In one or more dimensions, the squared exponential covariance, $K(t) = \exp(-t^2)$ which produces analytic realizations is also excluded since its spectral density, $\sqrt{\pi} \exp(-\omega^2/4)$, goes to zero too quickly to satisfy (3). The triangular covariance function $K(t) = (1 - |t|)^+$ which has documented undesirable interpolation properties is excluded as well since its spectral density, $2(1 - \cos(\omega))/\omega^2$, oscillates too violently to flatten out in the relative sense required by (3) (Stein 1999).

It should also be noted the condition in (3) is a more general condition than one posed in Stein (2002), wherein g is required to be asymptotically proportional to a regularly varying function. Like Stein (2011), Stein (2002) uses the screening effect to guide generation of restrictions on g . We use (3) over the condition in Stein (2002) because it describes a more general class of functions. Additional screening effect analysis can be found in Stein (2015), wherein conditions showing when a screening effect will not hold are given. Stein (2015) also supports use of models satisfying exactly the condition in (3).

Beyond half-spectral methods, there are a host of other techniques that can be used to produce valid non-separable space-time and other multidimensional covariance models. Brown et al. (2000) present a convolution based method to produce a flexible anisotropic class of models. Ma (2003) presents a mixture type method to combine marginal covariances into a nonseparable model. Kolovos et al. (2004) provide an overview of many methods including full spectral methods and methods requiring solution of stochastic partial differential equations. Furthermore, Kolovos et al. (2004) present a technique to produce models valid in many dimensions from one-dimensional models via weighted sums. Genton (2007) approximates non-separable models using separable covariances. Fuentes et al. (2007) and Stein (2005a) use full spectral representations to define non-separable models.

In this paper, we advance the half-spectral covariance model representation in (2) and show that only specific forms of functions \mathbb{C} , f , δ and θ will lead to models satisfying the condition in (3). In the next section of this paper, we show space-time models will satisfy (3) independent of the specification of θ as long as θ is locally bounded. This conclusion allows our remaining analysis in this section to focus on fully-symmetric models. We develop two restrictions on the set of functions, \mathbb{C} , f and δ and prove that models violating these restrictions will not meet (3). These restrictions are subsequently used in Section 3 to guide development of two new classes of half-spectral models. One class of models satisfies both restrictions presented in Section 2. The second class satisfies only one of the restrictions in Section 2. For both classes, we show how smoothness for a fully symmetric model in space and time may be changed by specification of components of \mathbb{C} , f and δ . In Section 4, we fit examples from both classes of models to daily average wind speed data recorded at different spatial locations across Ireland. We compare fits of these models to other models of similar form. We find models in the classes we develop to be easily adaptable to the data we consider, and we show they better capture space-time interaction in the Irish wind dataset compared to other separable and other non-separable models.

2 Restrictions on half-spectral forms

The half-spectral form in (2) in general does not meet the condition in (3). Let h be the spectral density of the spatial covariance function \mathbb{C} . The full spectral representation of K is

$$K(s, t) = \int_{\mathbb{R}} \int_{\mathbb{R}^d} f(\omega) \delta(\omega)^{-d} h\left(\frac{\lambda - \theta(\omega)\phi}{\delta(\omega)}\right) \exp(is^T \lambda + it\omega) d\lambda d\omega. \quad (4)$$

Let $g(\lambda, \omega)$ be the fully-symmetric spectrum, $g(\lambda, \omega) = f(\omega) \delta(\omega)^{-d} h(\lambda/\delta(\omega))$; therefore, we can write the general spectrum of K as $g(\lambda - \theta(\omega)\phi, \omega)$. To show when $g(\lambda, \omega)$ or $g(\lambda - \theta(\omega)\phi, \omega)$ satisfy (3), we first establish the result in Theorem 1 that states: $g(\lambda, \omega)$ satisfies (3) if and only if $g(\lambda - \theta(\omega)\phi, \omega)$ satisfies (3) as long as a mild regularity condition requiring $\theta(\omega)$ to be locally bounded is satisfied. This result permits the theoretical focus of this section to center on fully-symmetric models without much reducing the generality of our results. With Theorem 1 established, the remaining theory in this section gives necessary conditions and restrictions on a fully-symmetric model (defined only by f , \mathbb{C} and δ) that must be met in order for g to satisfy (3). Proofs of Theorem 1 and the Restrictions presented in this section are in the Appendix.

Theorem 1 *Let $g(\lambda, \omega)$ be the full spectrum of a space-time covariance function where λ is the spatial wavenumber and ω is the temporal frequency. Let $\theta(\cdot)$ be an odd function, and let ϕ be a unit vector in \mathbb{R}^d . The following two statements hold:*

- (i) *Let $\theta(\cdot)$ be locally bounded. The full spectrum $g(\lambda, \omega)$ satisfies (3) if and only if $g(\lambda - \theta(\omega)\phi, \omega)$ satisfies (3).*
- (ii) *Let $g(\lambda, \omega)$ be strictly positive and let $g(\lambda, \omega)$ satisfy (3). If $g(\lambda - \theta(\omega)\phi, \omega)$ satisfies (3), then $\theta(\omega)$ must be locally bounded.*

Remark Statement (i) is bolstered in a practical sense by statement (ii). Model building strategies often start with a fully-symmetric model that may be tweaked to be not fully symmetric. If we accept (3) as a desirable condition, both the fully-symmetric and the not fully-symmetric versions of a model should satisfy (3). Statement (ii) therefore establishes that the regularity condition in statement (i) (the local boundedness of $\theta(\omega)$) is a natural condition.

With Theorem 1 in mind, we focus on fully-symmetric half-spectra $f(\omega)\mathbb{C}(s\delta(\omega))$ with full spectra $g(\lambda, \omega)$. For $g(\lambda, \omega)$ to meet the condition in (3), we show the spectrum h and function δ must lead to a factorizable form: $h(\lambda/\delta(\omega)) = H(\lambda, \omega)\delta(\omega)^d/f(\omega)$, where H satisfies (3). Though h and δ of this form can be seen to trivially meet (3), the restrictions developed in the following paragraphs establish certain more flexible forms will not satisfy (3).

Since $g(\lambda, \omega) = f(\omega)\delta(\omega)^{-d}h(\lambda/\delta(\omega))$, we can more simply write g in terms of a marginal function in ω and a joint function in λ and ω . Let $g(\lambda, \omega) = p(\omega)q(\lambda, \omega)$. The following restrictions establish necessary conditions on p and q that must be met in order for g to satisfy the condition in (3).

Restriction 1 *Let the spectral representation of covariance function K have the form $g(\lambda, \omega) = p(\omega)q(\lambda, \omega)$. The following two statements hold for spectral densities with this parameterization.*

- (i) *If g and q satisfy the condition in (3), then p must be constant in ω .*
- (ii) *If g satisfies the condition in (3), then for every point ω_0 such that $p(\omega_0)$ is finite and positive, the marginal spectrum $g_{\omega_0}^*(\lambda) = g(\lambda, \omega_0)$ must meet the condition in (3).*

Remark The requirement in (i) that q follows the condition in (3) may be weakened. Our proof for (i) needs only that $\lim q(\lambda, \omega_0 + u_0)/q(\lambda, \omega_0) \neq 1/C$, a fact which follows from q satisfying (3). This condition however seems quite arbitrary, and we have opted to use the stricter condition.

The implication of Restriction 1 is simply that the space-time interaction function δ cannot be determined completely independently of \mathbb{C} and f if we wish to use models that follow the condition in (3). Of note is the fact that this restriction applies to several models proposed in Cressie & Huang (1999), Gneiting (2002) and Stein (2005b).

One may define the spatial interaction function as $\delta(\omega) = f(\omega)^{1/d}$ to escape the conditions of Restriction 1; however, $\delta(\omega)$ defined in this way decreases to zero for increasingly large ω since f must be integrable. This behavior of δ contradicts intuition—for example, considering (2), we see $\delta(\omega) \rightarrow 0$ implies coherences between multiple time series are highest at high frequencies. More formally, δ decreasing to zero or having a finite limit as ω approaches infinity will not generally lead to models satisfying the condition in (3).

Restriction 2 *Let $g(\lambda, \omega) = p(\omega)q(\lambda/\delta(\omega))$, where $p(\omega)$ is positive and finite for all sufficiently large $|\omega|$, q is a non-negative, continuous, integrable function that is not identically zero, and $\delta(\omega)$ is even with a well-defined limit as $|\omega| \rightarrow \infty$. If g satisfies (3), then $\lim_{|\omega| \rightarrow \infty} \delta(\omega) = \infty$.*

Remark A version of this Restriction exists concerning $\delta(\omega)$ without a well-defined limit as $|\omega| \rightarrow \infty$. If $\liminf_{|\omega| \rightarrow \infty} \delta(\omega)$ is finite, then under the additional condition that q satisfies (3) it can be proved that g cannot satisfy (3). We let q be more general and include the stronger condition on δ in this Restriction to match the class of models we develop in Section 3.2.

Restrictions 1 and 2 confine the form of δ . Implicitly, the form of h is also restricted. For half-spectral models satisfying (3), if δ must increase to infinity, $f(\omega)\delta(\omega)^{-d}$ cannot be constant, implying $h(\lambda/\delta(\omega))$ itself cannot satisfy (3). On the other hand, for fixed ω , $h(\lambda/\delta(\omega))$ must satisfy (3) in λ by part (ii) of Restriction 1. One way forward is to choose h and δ such that $h(\lambda/\delta(\omega))$ admits the representation, $h(\lambda/\delta(\omega)) = H(\lambda, \omega)\delta(\omega)^d/f(\omega)$, where H satisfies (3). A half-spectral model thus specified will satisfy (3). Finally, by Theorem 1, with any locally bounded θ , $h((\lambda - \theta(\omega)\phi)/\delta(\omega))$ specified in this way will also lead to models satisfying (3).

3 New classes of models

We build two new classes of fully-symmetric models; thus $\theta(\omega) = 0$ for each of the classes of models we present in this section. If $\theta(\omega)$ is non-zero, models with space-time asymmetry can be defined. Importantly, as we showed in Section 2, a zero or non-zero $\theta(\omega)$ does not affect whether or not a given model will satisfy (3) as long as θ is a locally bounded function.

3.1 A class of models satisfying the natural condition

Let $h(\lambda) = \phi(\alpha^2 + \|\lambda\|^2)^{-(\nu+d/2+1/2)}$, the Matérn spectral density with smoothness $\nu+1/2 > 0$, $\alpha > 0$ an inverse range parameter and $\phi > 0$, a variance parameter. For a given f , to ensure Restrictions 1 and 2 are not violated, choose $\delta(\omega) = f(\omega)^{-1/2\frac{1}{\nu+1/2}}$. Plugging into the space-time spectrum in (4), a new set of spectral densities can be written

$$g(\lambda, \omega) = \phi \left(\alpha^2 f(\omega)^{-1/(\nu+1/2)} + \|\lambda\|^2 \right)^{-(\nu+(d+1)/2)}. \quad (5)$$

If f , itself a spectral density, satisfies $f \sim c_f \omega^{-k}$ for some $k > 1$ and c_f a positive constant depending on f , then (5) describes a class of models that satisfies (3). We write covariance functions with this half-spectrum K_f to indicate the temporal spectrum f determines a specific model in this class.

We use a Matérn spectral density in space to form this class of models for three reasons. First, it is a natural model to analyze since spatial modeling is often carried out using a Matérn covariance or a modification of a Matérn covariance. Second, its form leads to easy comparison with the conditions of Restrictions 1 and 2. Third, the Matérn spectrum gives (5) some desirable Matérn-like properties. For example, as we show in the following paragraphs, the parameter ν , is directly related to the spatial smoothness of the resulting Gaussian process. The $1/2$ added to ν ensures this formulation extends the space-time Matérn covariance naturally. If $f(\omega)$ is a one-dimensional Matérn spectral density with smoothness ν , then $K_f(s, t)$ is simply a $d + 1$ dimensional Matérn covariance function with smoothness ν . Note, in this case, $\nu > 0$, since f must be integrable.

The half-spectral representation of (5) is given simply by plugging the Matérn covariance into the relevant portion of (2).

$$\begin{aligned} f(\omega)\mathbb{C}(s\delta(\omega)) &= f(\omega) \frac{\pi^{d/2}\phi}{2^{\nu-1/2}\Gamma(\nu + (d+1)/2)\alpha^{2\nu+1}} \times \\ &\quad (\alpha\|s\|f(\omega)^{-1/2\frac{1}{\nu+1/2}})^{\nu+1/2}\mathcal{K}_{\nu+1/2}\left(\alpha\|s\|f(\omega)^{-1/2\frac{1}{\nu+1/2}}\right) \end{aligned} \quad (6)$$

where $\mathcal{K}_{\nu+1/2}$ is the modified Bessel function of the second kind with argument $\nu+1/2$ (Olver et al. 2010). If we let ϕ absorb terms not depending on ω in (6) and write $\mathcal{M}_{\nu+1/2}(s) = \|s\|^{\nu+1/2}\mathcal{K}_{\nu+1/2}(\|s\|)$, (6) can be simplified to

$$f(\omega)\mathbb{C}(s\delta(\omega)) = \phi f(\omega)\mathcal{M}_{\nu+1/2}\left(\alpha\|s\|f(\omega)^{-1/2\frac{1}{\nu+1/2}}\right). \quad (7)$$

Beyond satisfying the condition in (3), mean-square differentiable properties of this class of models are readily found. For a space-time process, $Z(s, t)$, governed by stationary covariance $K(s, t)$, $Z(s, t)$ is m times mean-square differentiable in time if and only if $\frac{\partial^{2m}}{\partial t^{2m}}K(0, 0)$

exists (Stein 2005a). Following (Stein & Weiss 1971, p. 5), k -th derivatives of $K(0, 0)$ in the temporal direction are known to exist if the purely temporal spectral density, $f(\omega)$, is integrable and $\int f(\omega)|\omega|^k < \infty$. Hence, if f integrable and $f \sim c_f|\omega|^{-k}$ as ω approaches infinity, then $Z(s, t)$ will be m times mean-square differentiable in time if $(k - 1)/2 > m$. Similar statements hold regarding the mean-square differentiability of $Z(s, t)$ in space. To determine spatial mean-square differentiability, we mirror the argument in (Stein 2005a, p. 313), showing $Z(s, t)$ will be m time mean-square differentiable in space if and only if $(\nu + 1/2) \cdot (k - 1)/k > m$. Therefore, with these models, smoothnesses in space and time can be defined via ν and the limiting behavior of f determined by k . We call $(\nu + 1/2) \cdot (k - 1)/k$ and $(k - 1)/2$ the effective smoothnesses of the purely spatial and purely temporal processes respectively. Since $\nu > -1/2$ and $k > 1$, arbitrary effective smoothnesses in space and time can be specified using models in this class.

Example 1 (Matérn in Time)

Let $f(\omega) = (\beta^2 + \omega^2)^{-(\kappa+1/2)}$, a one dimensional Matérn spectral density. To avoid overparameterization, the scale parameter is fixed at one. Plugging f into (5) and (7), we get the spectral and half-spectral forms.

$$\begin{aligned} g(\lambda, \omega) &= \phi \left(\alpha^2 (\beta^2 + \omega^2)^{(\kappa+1/2)/(\nu+1/2)} + \lambda^2 \right)^{-(\nu+(d+1)/2)}; \\ f(\omega)\mathbb{C}(s\delta(\omega)) &= \phi \left(\beta^2 + \omega^2 \right)^{-(\kappa+1/2)} \mathcal{M}_{\nu+1/2} \left(\alpha \|s\| (\beta^2 + \omega^2)^{1/2 \frac{\kappa+1/2}{\nu+1/2}} \right). \end{aligned} \quad (8)$$

Here $k = 2\kappa + 1$; hence this process is $\lceil \kappa \frac{\nu+1/2}{\kappa+1/2} - 1 \rceil$ times mean-square differentiable in space and $\lceil \kappa - 1 \rceil$ times mean-square differentiable in time, where $\lceil \cdot \rceil$ is the ceiling function. Again, it is readily checked that if $\kappa = \nu$, K_f reduces to a $d + 1$ dimensional Matérn model with perhaps different range parameters in time and space, depending on β and α . A numerically calculated contour plot for $\nu = 1/8$, $\kappa = 2$ is in Figure 1(a). Marginal plots of $K_f(0, s)$ and $K_f(t, 0)$ are in Figure 1(b).

Of note in Figure 1(a) is the fact that a “dimple”-type property is present in this model (Kent et al. 2011). Though the dimple we see in Figure 1(a) may not strictly follow the dimple definition given in Kent et al. (2011), the general phenomenon—that a dimple is indicative of higher covariances existing at non-zero spatial lags when the temporal lag is nonzero—is clearly present in this model. A broader definition of a dimple may be the following: a dimple exists if a contour of a covariance function is a non-convex shape. Empirically, this appears to be caused by the large difference in κ and ν and $\nu < \kappa$. For κ and ν closer together or $\nu > \kappa$, numerical experiments suggest this dimple either disappears or gets pushed further from the origin. Since half-spectral forms in space can be defined similar to (8), models without apparent dimples with flexible smoothnesses in space and time can be defined. Note, this model is a special case of a class of models considered by Stein (2005a), but the half-spectral form of this model is not given in Stein (2005a) nor is the connection of this model to Matérn models mentioned therein.

Example 2 (Continuous AR(2) in Time)

From (Priestley 1983, p. 239), a continuous analog of an auto-regressive process of order 2 can be defined by the following two parameter spectrum.

$$f(\omega) = \frac{1}{\pi} \left(\frac{\beta_1 \beta_2}{(\beta_2 - \omega^2)^2 + \beta_1^2 \omega^2} \right).$$

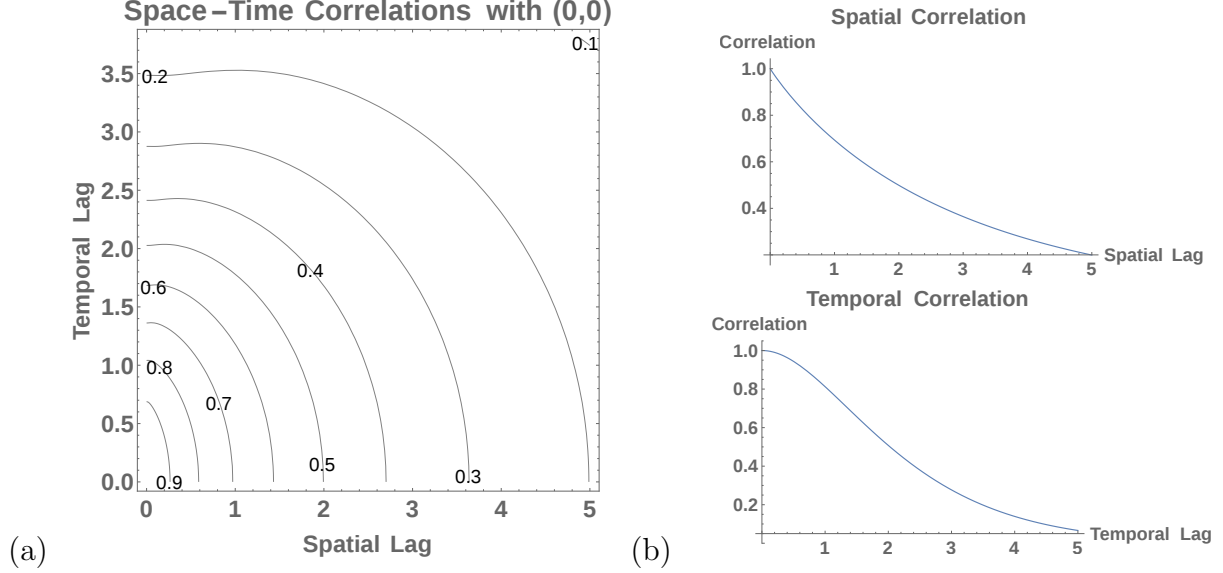


Figure 1: Contour and marginal covariance plots for the Matérn in Time model in Example 1. Effective smoothness in space is 0.5. Effective smoothness in time is 2.

For large ω , $f(\omega) \sim c_f \omega^{-4}$; hence this process is one time mean-square differentiable in time. Contour and marginal plots are in Figure 2.

The oscillatory behavior of an AR(2) process is documented in Figure 2(a) and (b) by the negative covariances. There is a hint of a dimple in Figure 2(a), though the effect does not appear to be severe. Like the Matérn in Time model in Example 1, the dimple effect becomes more pronounced when smoothness in space is much lower than smoothness in time, and it either disappears or becomes unnoticeable when smoothness in space is much larger than smoothness in time.

3.2 A class of half-spectral models satisfying Restriction 2

Restriction 2 effectively shows that a model of the form (2) must satisfy $\delta(\omega) \rightarrow \infty$ as $\omega \rightarrow \infty$ in order for (3) to hold. Of course, models may satisfy $\delta(\omega) \rightarrow \infty$ as $\omega \rightarrow \infty$ without satisfying (3). We consider $\delta \rightarrow \infty$ as $\omega \rightarrow \infty$ a condition of interest in its own right. In particular, it forces coherences between time series at distinct spatial sites to tend to 0 as $\omega \rightarrow \infty$. In this Section, we define a class of models satisfying $\delta(\omega) \rightarrow \infty$ as $\omega \rightarrow \infty$ but not (3).

Briefly, if f is a probability density (letting \mathbb{C} be more generally a covariance function), we may write

$$K(s, 0) = \mathbb{E}_\Omega [\mathbb{C}(s\delta(\Omega))] \quad (9)$$

where Ω is distributed according to f . Hence, $K(s, 0)$ depends only on the distribution of $\delta(\Omega)$. Note further that since $K(0, t)$ is determined completely by f , $K(0, t)$ depends solely on the distribution of Ω . Since f is even, Ω must be a symmetric random variable, and since $\delta(\omega)$ is even and positive, $\delta(\Omega)$ is distributed as the absolute value of a symmetric random variable. Let A and B be continuous, symmetric random variables. We can specify $\Omega \sim B$

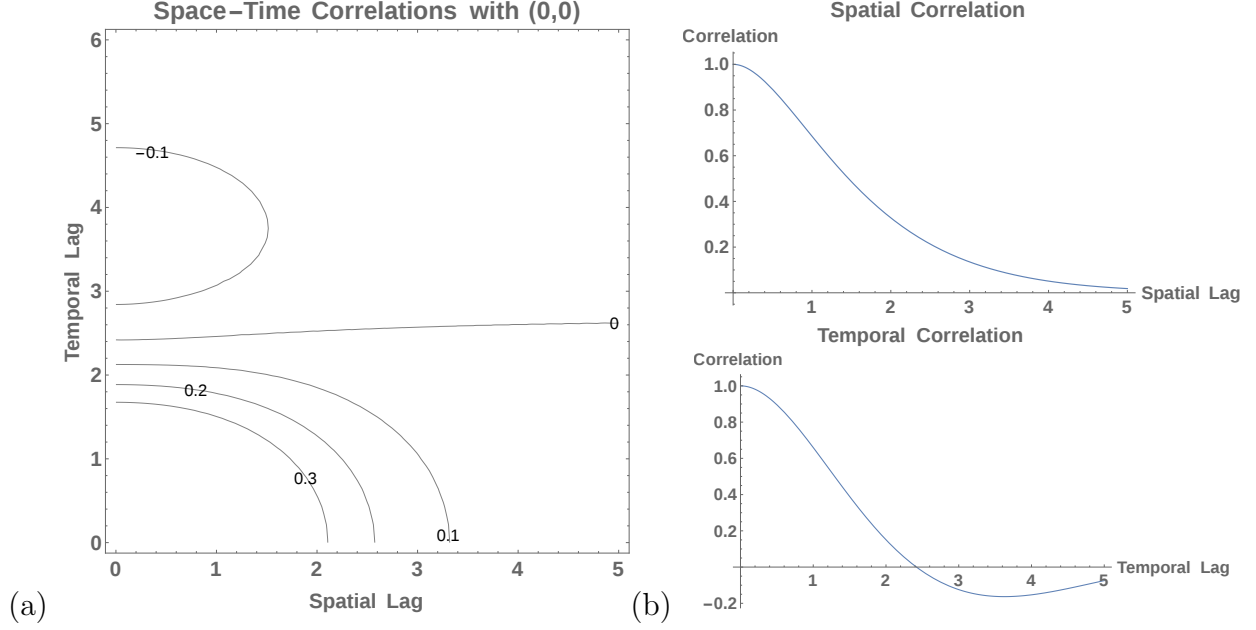


Figure 2: Contour and marginal covariance plots for the AR(2) in Time model in Example 2. Effective smoothnesses in space and time are each $3/2$. Auto-regressive parameters $\beta_1 = \beta_2 = 1$.

and $\delta(\Omega) \sim |A|$; thus, $\delta(\omega) = |F_A^{-1}(F_B(\omega))|$ for F_A and F_B distribution functions. Therefore, in principle, $K(s, 0)$ and $K(0, t)$ can be defined independently without using constant δ . Moreover, δ of this form increases unboundedly; hence this set of models does not violate Restriction 2.

Similar to a class of models defined in Ma (2003), we further define this class by letting $\mathbb{C}(s) = \exp(-\|s\|^2)$. With this \mathbb{C} , the spatial model in (9) is a class of functions studied in Schoenberg (1938). If we allow $|A|$ to follow an arbitrary positive probability distribution on \mathbb{R} , Schoenberg (1938) proved spatial models in (9) with \mathbb{C} squared exponential characterize all covariance functions that are valid in all dimensions. In other words, the class of models (9) can marginally produce is quite vast. Nonetheless, \mathbb{C} as squared exponential has a spectrum proportional to $\exp(-\|\lambda\|^2/4)$; therefore, by Restriction 1 (ii), these models will not meet the condition in (3).

Example 3 (Marginal Matérn)

Let $f = f_B = \beta^{2\kappa} \frac{\Gamma(\kappa+1/2)}{\sqrt{\pi}\Gamma(\kappa)} (\beta^2 + \omega^2)^{-(\kappa+1/2)}$ with $\kappa, \beta > 0$. Let $\mathbb{C}(s) = \phi \exp(-\alpha^2 \|s\|^2)$, and let A be defined symmetrically by $|A| \sim 1/\sqrt{2\chi_{2\nu}^2}$, where $\nu, \phi, \alpha > 0$. It can be shown via results in (Bateman 1954, p. 146) that $K(s, 0) = \phi/(2^{\nu-1}\Gamma(\nu))\mathcal{M}_\nu(\alpha s)$. Since f_B is a Matérn spectral density, $K(0, t) = \phi/(2^{\kappa-1}\Gamma(\kappa))\mathcal{M}_\kappa(\beta t)$. In Figure 3, we give marginal and contour plots comparable to those in Figure 1. Figure 3(a) shows the dimple in this model is much more pronounced than that in Example 1. Empirically, the dimple becomes less pronounced when the smoothnesses of the purely spatial and purely temporal processes are closer together, but the dimple does not disappear even when $\nu = \kappa$. Explicit expressions for the half-spectrum of this model are given in the Appendix.

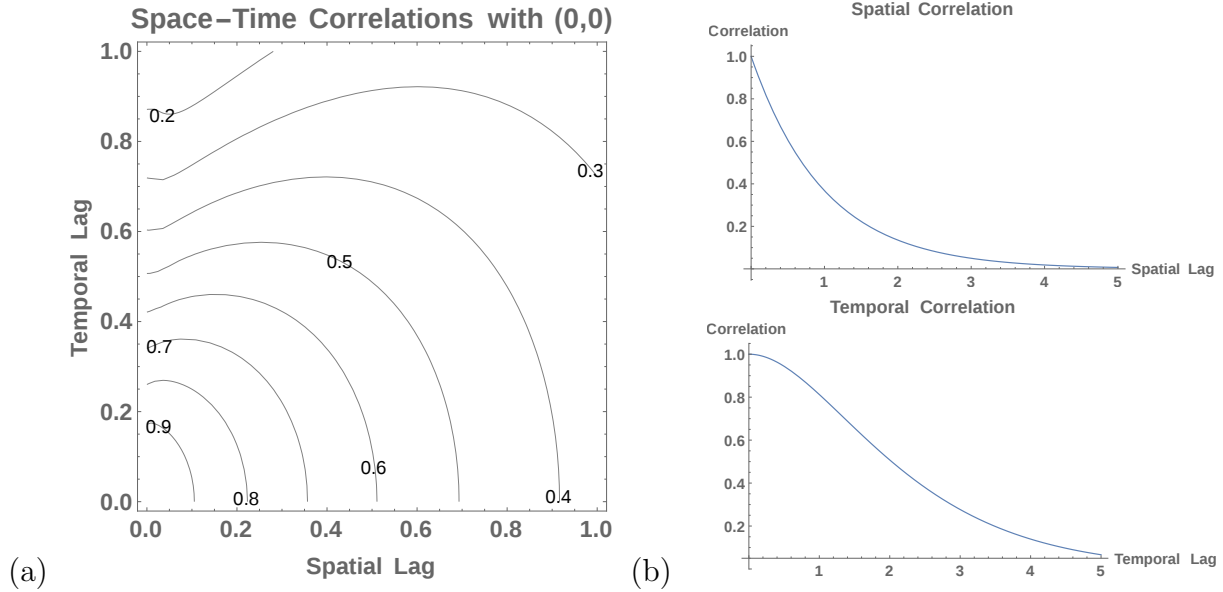


Figure 3: Contour and marginal covariance plots for the Marginal Matérn model. Effective smoothness in space is $1/2$. Effective smoothness in time is 2.

4 Application

We apply examples from the two classes of models presented in Section 3 to the Irish wind dataset first analyzed by Haslett & Raftery (1989). We compare fits from models in either class to similarly parameterized separable and non-separable models. In making these comparisons, we focus on space-time interaction properties of the models we fit; hence, when possible, marginal spatial and temporal covariances and smoothnesses are fixed at common values across the models we fit. We focus on space-time interaction to show the models we develop may be natural first choices to consider when analyzing space-time data.

The Irish wind dataset is a record of daily average wind speeds collected at 12 sites in Ireland from 1961 to 1978. This dataset has been analyzed several times previously by Haslett & Raftery (1989), Gneiting (2002), de Luna & Genton (2005) and Stein (2005b) among others. Like these authors, we deseasonalize the data by regressing wind-speed on a few (four) annual harmonics, and we discard data from the Rosslare site due to nonstationarities. We also take square-roots to obtain more marginally Gaussian data, and we subtract station means of these deseasonalized and transformed data. Finally, we divide each station's time series by its respective sample standard deviation. The resulting dataset contains $6,574 \times 11 = 72,314$ observations with none missing.

There are subtleties in these data. Most notably these data have been documented as having longer memory in time than covariance models with exponential tails can produce (Haslett & Raftery 1989). Therefore, to make best comparisons within groups of models, we separate comparisons based on the temporal tail behavior (exponential tails versus algebraic tails) of the covariance models we consider.

The wind process also has properties that are not fully-symmetric (Gneiting 2002). The primary model comparisons we make in this section involve fully-symmetric models because

these provide the best comparisons to established models. However, at the end of the application we fit a model that does not exhibit full-symmetry by using a simple model for the phase shift present in (2).

4.1 Regular monitoring data and model fitting

Let $Z(s, t)$ be the daily average wind speed at a spatial location s in Ireland observed on day t . For these observations, we define models in 4-dimensional Euclidean space, $(s, t) \in \mathbb{R}^3 \times \mathbb{R}$, that we restrict to the actual domain, the sphere \times time. Since $Z(s, t)$ is observed at each spatial site, s_i , for each time point, j , we can write observational points in the form (s_i, j) for all $i \in \{1, \dots, 11\}$ and $j \in \{1, \dots, 6574\}$. Data of this form is called regular monitoring data (Stein 2005b).

Fitting the models developed in Section 3 to regular monitoring data can be done by either computing covariances via a numerical evaluation of the integral in (2) or by viewing regular monitoring data as a multivariate time-series and using spectral methods. We focus on spectral fitting in this paper; however, to quickly evaluate the integral in (2) at multiple spatial lags for a fixed temporal lag, discrete Fourier transforms can be used to speed computations. We primarily use spectral fitting for convenience. It should also be noted that $\mathbb{C}(s\delta(\omega))$ can be seen as the cross-spectra of spatially situated multivariate time series (Stein 2005b).

Spectral fitting requires maximization of a multivariate Whittle likelihood. The models in Section 3 are defined in continuous space and time; hence, if we wish to use the multivariate Whittle likelihood, which estimates the spectrum of a discretely observed time series, we will need to account for aliasing. We use a simple aliasing correction. For the continuous model $f(\omega)\mathbb{C}(s\delta(\omega))$, we approximate a finite version of this function with $\sum_{j=-m}^m f(\omega + 2j\pi)\mathbb{C}(s\delta(\omega + 2j\pi))$ for $-\pi < \omega < \pi$. Note this sum is a truncation of the proper correction in which the sum goes from $-\infty$ to ∞ . We pick the truncation parameter $m = 50$ for convenience. For the models we fit, increasing m beyond 50 appeared to have little effect on likelihoods and parameter estimates; however, in general, tail behavior of spectrums can vary tremendously; hence, this truncation should be determined carefully in specific applications.

4.2 Results

4.2.1 Models with Exponentially Decaying Temporal Tails

We fit the models in Examples 1 and 3 to these data. For comparison, we fit a Matérn model as well as a separable model with exponential marginal covariances. We also fit a model in Cressie & Huang (1999) that was found to best fit a different dataset of wind speeds. Since the best-fitting model in Cressie & Huang (1999) is separable with squared exponential covariance in space and exponential covariance in time, the half-spectral form of this model is straightforward to obtain. For the Cressie & Huang (1999) model, $f(\omega)\mathbb{C}(s\delta(\omega)) = \phi(\beta^2 + \omega^2)^{-1} \exp(-\alpha^2 \|s\|^2)$. Note δ is constant in this model due to its separability. Cressie & Huang (1999) fit their models using a weighted least squares method to match the model variogram to the empirical variogram. In the model variogram they include a term related

Table 1: Comparison of log-likelihoods of short memory in time models fitted to the Irish wind dataset.

Whittle Log-Likelihoods of Models with Exponential Temporal Tails			
	Log-Likelihood	Diff. From Ex. 1	# Param. Fit
Example 1 ($\kappa = 1/2$)	20,318	0	5
Matérn (Example 1, $\kappa = \nu = 1/2$)	20,199	119	4
Example 3 ($\kappa = 1/2$)	19,441	877	5
Separable Exponential	18,703	1,615	4
Cressie & Huang (1999)	18,378	1,940	4

to $\|s\|$. Since we subtract site means prior to analysis of the Irish wind data, we do not need to use an analog of this term in our models.

The only modification we make to all four exponential temporal tail models is to add a spatial nugget similar to Gneiting (2002) and Stein (2005*b*). The half-spectral representations of our models therefore take the general form

$$f(\omega) [\mathbb{C}(s\delta(\omega)) + \eta^2 \mathbb{I}_{\{s=0\}}] \quad (10)$$

where $\mathbb{I}_{\{\cdot\}}$ is an indicator function and $\eta^2 \geq 0$.

For models with flexibility in determining smoothness in time (Examples 1 and 2 and the Matérn models), temporal smoothness parameters, κ , were fixed at $1/2$ to allow for better comparison of all models. Note for the Matérn model, this also fixes the spatial smoothness. To evaluate the assumption that $\kappa = 1/2$, we fit a univariate Matérn model to a subset of the time series at each station, and we found $\kappa = 1/2$ to be a reasonable estimate of temporal smoothness.

A comparison of the multivariate Whittle likelihoods is in Table 1. Here, the model from Example 1 is shown to provide the best fit for these data. That Example 1 outperforms a Matérn model is expected since Example 1 extends the Matérn class. The added flexibility in specifying the spatial smoothness appears to add a small amount to the log-likelihood in this case. With Example 1, the effective smoothness in space is estimated to be 0.4; hence a modest improvement in log-likelihood is not surprising. For processes estimated to have drastically different smoothnesses in space and time, we may expect a larger improvement in using Example 1 over a Matérn model.

A look at the remaining fits shows smoothness and space-time interaction matter in terms of model fitting. For example, the Matérn model compared to the separable exponential model shows an increase in 1,496 of the log-likelihood. Both models have the same number of fitted parameters and the same smoothnesses in space and time; yet, the Matérn appears able to capture the space-time interaction of the Irish wind process better. Comparing the separable exponential to the model in Cressie & Huang (1999), we see modeling variations in space as analytic is perhaps unreasonable. Comparing Example 3 to Example 1, we also see simply being able to fit marginal covariances may not directly lead to good models. Nonetheless, Example 3 appears to capture space-time interaction in these data better than the separable models present in Table 1.

Table 2: Comparison of log-likelihoods of longer memory models fitted to the Irish wind dataset.

Whittle Log-Likelihoods of Models with Algebraic Temporal Tails				
		Log-Likelihood	Difference From K_{f_G}	# Parameters Fit
K_{f_G}	$(\nu = 1/2)$	21,655	0	4
G	$(\gamma = 0, \kappa = 1)$	21,245	410	4

4.2.2 Models with Algebraic Temporal Tails

Gneiting (2002) uses the following covariance model, denoted by G , to analyze these data:

$$G(s, t) = \phi(\beta|t|^\kappa + 1)^{-1} \exp\left(-\frac{\alpha\|s\|}{(\beta|t|^\kappa + 1)^{\gamma/2}}\right) + \eta^2(\beta|t|^\kappa + 1)^{-1}\mathbb{I}_{\{s=0\}} \quad (11)$$

where $\beta > 0$, $\kappa \in [0, 2]$ and $\gamma \in [0, 1]$. The last term in (11) is the spatial nugget as also seen in (10). The temporal marginal covariance $G(0, t)$ has algebraic tail behavior. When $\kappa \leq 1$, G is a classical long memory model in time. In the case where $\kappa > 1$, $G(s, t)$ is not formally long memory in time since $G(0, t)$ will be integrable; however, the algebraic decay in $G(0, t)$ for large t is much slower than the exponential decay present in the models in Table 1 and fits the Irish wind data better than exponentially decaying models in Section 4.2.1.

To fit G using a multivariate Whittle likelihood method, the half-spectrum of G is needed, but we do not know of a general closed form half-spectral representation of G currently. However, specific separable half-spectral forms of G can be found. For $\gamma = 0$, G is separable and exponential in space. For $\kappa = 1$, a closed form for the spectral density of $G(0, t)$ is readily calculated (see Appendix); hence, the half-spectrum of a separable G that is long-memory in time can be specified as long as $\kappa = 1$ and $\gamma = 0$ are fixed. Note that Gneiting (2002) estimated $\kappa = 1.544$. If we instead let $\kappa = 2$, a closed form spectrum for $G(0, t)$ can also be found, but it is exponential and thus $G(0, t)$ for this model will not meet the condition in (3). With $\kappa = 1$, $G(0, t)$ meets the condition in (3) since the spectrum of $G(0, t)$ denoted f_G decays proportional to ω^{-2} (details in the Appendix).

Since f_G decays asymptotically proportional to ω^{-2} , we can plug f_G into the class in (7) to obtain K_{f_G} , a model that satisfies the condition in (3). Fitting both K_{f_G} with a spatial nugget and G with $\kappa = 1$ and $\gamma = 0$ by their respective half-spectra, we obtain the results in Table 2. We fix $\nu = 1/2$ for K_{f_G} to achieve the same smoothness in space for K_{f_G} as exists for G with $\gamma = 0$ and $\kappa = 1$.

Table 2 shows again that compared to similar non-separable models, a separable model will not adequately fit these data. This exercise shows further that the class of models presented in Section 3.1 is quite flexible and can be adapted straightforwardly to many different scenarios. Comparing the log-likelihoods of the short memory models in Table 1 to the longer memory models in Table 2, we find that G with $\gamma = 0$, a separable model, to better fit these data than the non-separable models in Table 1. The difference between G in Table 2 and Example 1 from Table 1 is 927 Whittle log-likelihood units. This is not completely surprising given that long memory properties of these data are apparent in looking at marginal correlation plots (Haslett & Raftery 1989).

Table 3: Comparison of log-likelihoods of longer memory models fitted to the Irish wind dataset. In contrast to Table 2, G is a non-separable model.

Space-Time Log-Likelihoods of Models with Algebraic Temporal Tails				
		Log-Likelihood	Difference From K_{f_G}	# Parameters Fit
K_{f_G}	$(\nu = 1/2)$	21,471	0	4
G	$(\kappa = 1)$	21,327	144	5

As a final comparison, we fit K_{f_G} and a non-separable version of G by maximizing space-time likelihoods (as opposed to the frequency domain Whittle likelihoods). The size of this dataset (72,314 observations) is quite difficult for space-time likelihood fitting of these models without making use of computational shortcuts. Specifically, for these data, calculating a log-determinant and a solve involving a $72,314 \times 72,314$ covariance matrix is required to compute the space-time likelihood under the multivariate Gaussian assumption; however, even storing a matrix of this size in RAM will not be possible on most machines as it requires over 40GB of memory using standard precisions. To perform this calculation, we make use of the regular monitoring structure of these data and the temporal stationarity of these models. Sorting the data first by time and second by location leads to a Block-Toeplitz covariance matrix with blocks of size 11×11 . Using a Block-Toeplitz Levinson-type algorithm, one can compute the log-determinant and solve terms simultaneously with greatly reduced computational burden (Akaike 1973).

The covariance model G is given in its space-time representation in (11). The space-time covariance function K_{f_G} must be computed numerically from the half-spectrum specified by (6). We approximate the covariance function K_{f_G} via a finite approximation of the integral in (2). As mentioned in Section 4.1, an approximation of the integral in (2) for many different time lags, t , can be computed quickly using the fast Fourier transform.

To ensure G and K_{f_G} are most comparable, we fix $\nu = 1/2$ for K_{f_G} and $\kappa = 1$ for G . With these parameter values, both models have the same smoothnesses in space and (up to range and scale parameters) identical marginal temporal covariance models. For G , the separability parameter γ is estimated from the data. Results from fitting these models are in Table 3.

Model K_{f_G} fits the data better than G while simultaneously requiring fewer estimated parameters. The separability parameter γ is estimated to be 0.5; hence G is non-separable. The marginal spatial and marginal temporal covariances in both models are quite similar; thus the slight edge K_{f_G} appears to have over G in this case may be due to the ability of K_{f_G} to better capture space-time interactions in these data. Comparing Table 3 to Table 2, we see a difference in likelihoods for K_{f_G} . Due to the approximations made in calculating both Whittle-type and space-time likelihoods, some discrepancy is expected.

4.2.3 A Space-time Asymmetric Model

Starting with K_{f_G} , we generate a space-time asymmetric model, \tilde{K}_{f_G} , by picking $\theta(\omega)\phi$ to be non-zero. We use a simple linear translation model for the phase shift, giving $\theta(\omega) = \rho\omega$, where $\rho \in \mathbb{R}$. The spatial direction of this shift is determined by ϕ . Since $\theta(\omega)$ is locally

bounded, \tilde{K}_{f_G} will satisfy (3). Gneiting (2002) remarks on the east-west asymmetry in these data; hence, for simplicity, we fix $\phi \in \mathbb{R}^3$ to be a unit vector pointing in a direction consistent with an east-to-west direction over Ireland. In principle, ϕ can be estimated from the data. Lastly, for best comparison to K_{f_G} in Table 2, ν is again fixed at $1/2$ for \tilde{K}_{f_G} .

Using Whittle likelihood fitting, \tilde{K}_{f_G} has a log-likelihood of 21,922. Comparing to K_{f_G} in Table 2, we see a marked advantage to taking into account space-time asymmetry with these data. The increase in Whittle log-likelihood is 267 units. The value ρ was estimated to be positive; hence, \tilde{K}_{f_G} is consistent with the simple measures of asymmetry used in Gneiting (2002).

Lastly, it is important to note that in the context of modeling a process on the entire globe, this phase shift model will not be entirely sensible. The vector $\rho\omega\phi$ indicates a linear phase shift through 3-dimensional Euclidean space. Our data are on the globe however; therefore, an east-west phase shift is not everywhere accurately described by the single vector ϕ . For example, while ϕ is an east-to-west vector over Ireland, it is a west-to-east vector on the opposite side of the globe (for example, in Japan). A better way to model phase shifts on the globe in the context of half-spectral models is given in Stein (2005b), wherein phase shifts are akin to rotations of the globe. In this type of phase shift model, phases depend on differences in longitude and/or latitude and not a lag vector in \mathbb{R}^3 . Unfortunately, rotation based phase models do not fit exactly within the theory presented in Section 2. Therefore, we have opted to use the locally accurate phase shift described by $\rho\omega\phi$. Since Ireland covers only a small range of longitudes, using this simplification should not affect results in any substantial way.

5 Discussion

We developed two new classes of continuous space-time dependency models via half-spectra defined by the form $f(\omega)\mathbb{C}(s\delta(\omega))\exp(i\theta(\omega)\phi)$. Both classes of models allow for substantial flexibility in modeling the marginal processes defined by $K(s, 0)$ and $K(0, t)$. Moreover, both classes of models are non-separable. These models are further theoretically validated by meeting at least one of two restrictions we developed in Section 2 to guide our model building. These restrictions were developed through consideration of a natural condition posed in Stein (2011). Our model building focused on fully-symmetric models defined by $f(\omega)\mathbb{C}(s\delta(\omega))$; however, Theorem 1 in Section 2 established that inducing space-time asymmetry in these models using a non-zero $\theta(\omega)$ can be done naturally and easily.

We compared fully-symmetric examples from the new classes of models in this paper to separable models and to a non-separable model developed in Gneiting (2002) by fitting these models to the Irish wind dataset in Section 4. Models from the class of half-spectra defined in (6) were shown to fit the Irish wind data better than all other models we examine in this paper in terms of likelihood. Haslett & Raftery (1989) noted these data exhibit long-memory properties in time, and in Section 4, we showed the flexibility of the first class of models by fitting a model that is long-memory in time that also meets the condition (3). We also fit an adaptation of this model using a non-zero $\theta(\omega)$ to illustrate the space-time asymmetry present in these data. Half-spectral models of the form in (7) therefore appear to be quite flexible and may be a good starting point for many model building purposes.

Space-time asymmetry was modeled as a phase shift using $\theta(\omega)$ in this paper. Other methods exist that produce space-time asymmetry that may be of useful in half-spectral modeling. Geometric transformations of space in particular may be used. In such models, a spatial lag vector, s , may be substituted for a spatial lag shifted through time, $s + Vt$, where V is a vector indicating the magnitude and direction of the spatial shift. The interpretation of this asymmetry is straightforward. A model with this type of spatial shift is appropriate for physically moving processes. An added benefit to using this type of asymmetry is that an adaptation of Theorem 1 exists in this setting: these types of geometric transformations will not affect whether or not a model satisfies (3). Additional methods to produce space-time asymmetry may also be used in the half-spectral setting. Taking derivatives of a process in the manner of Jun & Stein (2007) can lead to space-time asymmetries as well as spatial anisotropy. Non-parametric deformation as seen in Sampson & Guttorp (1992) can also produce asymmetry.

The half-spectral forms present in this paper serve as one step toward a more coherent melding of spatial and temporal methods in statistics. A primary avenue of future research may be to develop computationally efficient fitting procedures for half-spectral models on data that do not have the nice structure of non-missing regular monitoring data. Adaptation of composite likelihood methods [Lindsay (1988), Vecchia (1988) and Stein et al. (2004)] may serve as one way forward as these methods allow for considerations of only subsets of the full dataset.

6 Appendix

In this appendix, we prove the Theorems and Restrictions presented in Section 2. We give explicit forms for the half-spectrum of the model from Example 3 and of the marginal temporal spectrum, f_G of G with parameter κ fixed at 1. We also give details on the fitting procedures used in Section 4.

6.1 Proofs of Theorem 1 Statement (i) and the Restrictions

Proofs of Theorem 1 Statement (i) and Restrictions 1 and 2 can be found here. Proof of Theorem 1 Statement (ii) has been put in Section 6.2 because its proof relies on two lemmas.

Theorem 1, Statement (i) 1 *Let $g(\lambda, \omega)$ be the full spectrum of a space-time covariance function where λ is the spatial wavenumber and ω is the temporal frequency. Let $\theta(\cdot)$ be an odd function, and let ϕ be a unit vector in \mathbb{R}^d . The following statement holds:*

- (i) *Let $\theta(\cdot)$ be locally bounded. The full spectrum $g(\lambda, \omega)$ satisfies (3) if and only if $g(\lambda - \theta(\omega)\phi, \omega)$ satisfies (3).*

Proof First, suppose $g(\lambda, \omega)$ satisfies (3). If we can prove $\|(\lambda - \theta(\omega)\phi, \omega)\| \rightarrow \infty$ as $\|(\lambda, \omega)\| \rightarrow \infty$, we will have proved $g(\lambda - \theta(\omega)\phi, \omega)$ will satisfy (3). When $\|(\lambda, \omega)\| \rightarrow \infty$, there are two cases to examine: $|\omega| \rightarrow \infty$ and the case where ω is bounded as $\|\lambda\|$ diverges.

If $|\omega| \rightarrow \infty$, $\|(\lambda - \theta(\omega)\phi, \omega)\|$ diverges to infinity. The remaining case, where ω is bounded as λ grows, must be considered. Since θ is locally bounded and ω is bounded, there exists some fixed A such that $|\theta(\omega)| < A$ for all ω as $\|(\lambda, \omega)\| \rightarrow \infty$. We have

$$\begin{aligned}\|\lambda - \theta(\omega)\phi\| &= \sqrt{\lambda^T \lambda - 2\theta(\omega)\lambda^T \phi + \theta(\omega)^2} \\ &\geq \sqrt{\|\lambda\|^2 - 2\|\lambda\|A}.\end{aligned}$$

Thus, $\|\lambda\| \rightarrow \infty$ implies $\|\lambda - \theta(\omega)\phi\| \rightarrow \infty$.

The reverse direction may be proved by recognizing $g(\lambda, \omega)$ as simply a phase shifted version of $g(\lambda - \theta(\omega)\phi, \omega)$ using the valid phase shift function $-\theta(\omega)$.

Restriction 1 *Let the spectral representation of covariance function K have the form $g(\lambda, \omega) = p(\omega)q(\lambda, \omega)$. The following two statements hold for spectral densities with this parameterization.*

- (i) *If g and q satisfy the condition in (3), then p must be constant in ω .*
- (ii) *If g satisfies the condition in (3), then for every point ω_0 such that $p(\omega_0)$ is finite and positive, the marginal spectrum $g_{\omega_0}^*(\lambda) = g(\lambda, \omega_0)$ must meet the condition in (3).*

Proof For both (i) and (ii), we prove the contrapositive of each statement. First we prove (i). Assume p is not constant and q satisfies (3). Then there exists some ω_1 and u_1 and R such that $p(\omega_1 + u_1)/p(\omega_1) = C > 1$ with $|u_1| < R$. Let ω be fixed at ω_1 ; hence $\|\lambda\| \rightarrow \infty$ in the limit in (3). Taking the supremum in (3), we have

$$\begin{aligned}\sup_{\|(u,v)\| < R} \left| \frac{p(\omega_1 + u)q(\lambda + v, \omega_1 + u)}{p(\omega_1)q(\lambda, \omega_1)} - 1 \right| &\geq \left| \frac{p(\omega_1 + u_1)q(\lambda, \omega_1 + u_1)}{p(\omega_1)q(\lambda, \omega_1)} - 1 \right| \\ &= C \left| \frac{q(\lambda, \omega_1 + u_1)}{q(\lambda, \omega_1)} - \frac{1}{C} \right|\end{aligned}$$

From here, we can bound the limit

$$\begin{aligned}\lim_{\|\lambda\| \rightarrow \infty} \sup_{\|(u,v)\| < R} \left| \frac{p(\omega_1 + u)q(\lambda + v, \omega_1 + u)}{p(\omega_1)q(\lambda, \omega_1)} - 1 \right| &\geq C \lim_{\|\lambda\| \rightarrow \infty} \left| \frac{q(\lambda, \omega_1 + u_1)}{q(\lambda, \omega_1)} - \frac{1}{C} \right| \\ &= C - 1\end{aligned}$$

The last expression is strictly larger than 0; hence g cannot satisfy the condition in (3). By contraposition (i) is proved.

For case (ii), assume there exists ω_0 such that $p(\omega_0)$ is finite and positive, and $q(\lambda, \omega_0)$ does not meet (3) marginally in λ . Fix $\omega = \omega_0$, and let $u = 0$. Since $0 < p(\omega_0) < \infty$, $p(\omega_0)/p(\omega_0) = 1$. The condition in (3) on g is therefore equivalent to the same condition on $q(\lambda, \omega_0)$, which is not satisfied by assumption. Statement (ii) follows again by contraposition.

Restriction 2 *Let $g(\lambda, \omega) = p(\omega)q(\lambda/\delta(\omega))$, where $p(\omega)$ is positive and finite for all sufficiently large $|\omega|$, q is a non-negative, continuous, integrable function that is not identically zero, and $\delta(\omega)$ is even with a well-defined limit as $|\omega| \rightarrow \infty$. If g satisfies (3), then $\lim_{|\omega| \rightarrow \infty} \delta(\omega) = \infty$.*

Proof Like Restriction 1, we prove Restriction 2 by contraposition. Assume $\lim_{|\omega| \rightarrow \infty} \delta(\omega) = E$, where $E < \infty$. Take $\|(\omega, \lambda)\| \rightarrow \infty$ along a path for which $|\omega| \rightarrow \infty$ and $\lambda = \delta(\omega)v_0$, where $v_0 \in \mathbb{R}^d$ and $q(v_0) > 0$. Since q is integrable, there exists some $v_1 \neq 0$ and some R such that $q(v_0 + v_1/E) \neq q(v_0)$ and $\|v_1\| < R$. The supremum in (3) can be bounded from below:

$$\begin{aligned} \sup_{\|(u,v)\| < R} \left| \frac{p(\omega + u)q\left(\frac{1}{\delta(\omega+u)}(\delta(\omega)v_0 + v)\right)}{p(\omega)q(v_0)} - 1 \right| &\geq \sup_{\|(0,v)\| < R} \left| \frac{q\left(v_0 + \frac{v}{\delta(\omega)}\right)}{q(v_0)} - 1 \right| \\ &\geq \left| \frac{q\left(v_0 + \frac{v_1}{\delta(\omega)}\right)}{q(v_0)} - 1 \right| \end{aligned}$$

Since q is continuous and $\lim_{|\omega| \rightarrow \infty} \delta(\omega) = E$, the limit of the supremum in (3) is bounded away from zero: $\lim_{\|(\lambda,\omega)\| \rightarrow \infty} \sup_{\|(u,v)\| < R} |g(\lambda + v, \omega + u)/g(\lambda, \omega) - 1| > 0$.

6.2 Proof of Theorem 1, Statement (ii)

The proof of Theorem 1, Statement (ii) relies on the following two lemmas.

Lemma 1 *Let $g(\xi)$, $\xi \in \mathbb{R}^{d+1}$ be a non-negative, integrable function. If $g(\xi)$ satisfies (3), then $\lim_{\|\xi\| \rightarrow \infty} g(\xi) = 0$.*

Proof We prove this statement using contradiction. Suppose $\limsup_{x \rightarrow \infty} g(\xi_0(x)) \neq 0$ for a certain path indexed by a parameter, x , $\xi_0(x) : \mathbb{R}^+ \rightarrow \mathbb{R}^{d+1}$ such that $\|\xi_0(x)\| \rightarrow \infty$ as x grows large. Therefore, there exists some $\varepsilon > 0$ such that for each $i \in \{1, 2, \dots\}$, there exist some $x_i \geq i$ where $g(\xi_0(x_i)) > \varepsilon$. Since $\|\xi_0(x_i)\| \rightarrow \infty$ as $i \rightarrow \infty$, we can select a subsequence, written y_i for $i \in \{1, 2, \dots\}$ such that $\|\xi_0(y_i) - \xi_0(y_j)\| > 2$ for $i \neq j$.

Now consider $\sum_{i=1}^{\infty} \int_{B(\xi_0(y_i), 1)} g(\xi) d\xi$, where $B(\xi_0(y_i), 1)$ indicates the ball of radius 1 centered at point $\xi_0(y_i)$. Because the radius of these balls is less than or equal to half the distances between each center point, $\xi_0(y_i)$, these balls do not overlap. Thus the sum of their volume will lower bound the full integral $\int_{\mathbb{R}^{d+1}} g(\xi) d\xi$. To complete this proof by contradiction, we show $\sum_{i=1}^{\infty} \int_{B(\xi_0(y_i), 1)} g(\xi) d\xi$ explodes by showing terms at sufficiently large values of i will have a positive number as a lower bound.

We have $g(\xi_0(y_i)) > \varepsilon > 0$ for each i . And we have $g(\xi)$ satisfies (3). For a radius, R , set $R = 1$, and pick $\delta = 1/2$. Since $g(\xi)$ satisfies (3), it holds that there exists an $I \in \mathbb{R}$ such that for all $i > I$,

$$\frac{1}{2} < \frac{g(\xi_0(y_i) + U)}{g(\xi_0(y_i))} < \frac{3}{2}$$

for all points such that $\|U\| < 1$. Since $g(\xi_0(y_i)) > \varepsilon$, we find a lower bound for all points in the unit ball centered at $\xi_0(y_i)$, $\inf_{\xi \in B(\xi_0(y_i), 1)} g(\xi) \geq \varepsilon/2$. Therefore, $\int_{B(\xi_0(y_i), 1)} g(\xi) d\xi \geq \varepsilon \pi^{(d+1)/2} / ((d+1)\Gamma((d+1)/2))$ for each $i > I$. Thus $\int_{\mathbb{R}^{d+1}} g(\xi) d\xi$ diverges, but this contradicts the integrability assumption.

Lemma 2 Let $g(\xi)$, $\xi \in \mathbb{R}^{d+1}$, be a strictly positive, integrable function. Let $\eta(\cdot)$, $\xi_0(\cdot)$ and $U_0(\cdot)$ be a triplet, $(\eta(\cdot), \xi_0(\cdot), U_0(\cdot))$, where $\eta : \mathbb{R}^{d+1} \rightarrow \mathbb{R}^{d+1}$, $\xi_0 : \mathbb{R} \rightarrow \mathbb{R}^{d+1}$ and $U_0 : \mathbb{R} \rightarrow \mathbb{R}^{d+1}$ such that $(\eta(\cdot), \xi_0(\cdot), U_0(\cdot))$ has the following properties:

(i) $\|\xi_0(x)\| \rightarrow \infty$ as $x \rightarrow \infty$.

(ii) $\|\eta(\xi_0(x))\|$ is bounded.

(iii) For any $C > 0$ and for all $x \in \mathbb{R}^+$, there exists a $Y \in \mathbb{R}$ dependent only on C such that for all $y > Y$,

$$\|\eta(\xi_0(x) + U_0(y))\| > C.$$

If $g(\xi)$ satisfies (3), then the function $f(\xi) = g(\eta(\xi))$ cannot satisfy (3).

Proof Loosely, the properties of the triplet, $(\eta(\cdot), \xi_0(\cdot), U_0(\cdot))$, ensure that we can pick a point, y , large enough such that $\eta(\xi_0(x) + U_0(y))$ is sufficiently far from $\eta(\xi_0(x))$ for all x . This, together with Lemma 1, prevents (3) from being satisfied for $f(\xi)$.

More formally, consider the supremum in condition (3) for $f(\xi_0(x))$ and for any $R \in \mathbb{R}$.

$$\sup_{\|U\| < R} \left| \frac{f(\xi_0(x) + U)}{f(\xi_0(x))} - 1 \right| \geq \left| \frac{f(\xi_0(x) + U_0(y))}{f(\xi_0(x))} - 1 \right|$$

for any y such that $\|U_0(y)\| < R$. Since R can be arbitrarily large, we know this condition can be satisfied. Substituting in g on the right hand side, we find

$$\sup_{\|U\| < R} \left| \frac{f(\xi_0(x) + U)}{f(\xi_0(x))} - 1 \right| \geq \left| \frac{g(\eta(\xi_0(x) + U_0(y)))}{g(\eta(\xi_0(x)))} - 1 \right|.$$

We know $\|\eta(\xi_0(x))\|$ is bounded and that g is strictly positive; hence, there exists some $\varepsilon > 0$ such that $g(\eta(\xi_0(x))) > \varepsilon$ for all x .

By Lemma 1 and property (iii) we also know y (and thus R) can be chosen large enough such that $g(\eta(\xi_0(x) + U_0(y))) < \varepsilon/2$ for all x . The limit of $\sup_{\|U\| < R} \left| \frac{f(\xi_0(x) + U)}{f(\xi_0(x))} - 1 \right|$ as $x \rightarrow \infty$ must therefore be bounded below by $1/2$ and thus cannot equal 0. Since $\|\xi_0(x)\| \rightarrow \infty$ as $x \rightarrow \infty$, we have found a path, $\xi_0(x)$, that violates condition (3) for $f(\xi)$.

Theorem 1, Statement (ii) 1 Let $g(\lambda, \omega)$ be the full spectrum of a space-time covariance function where λ is the spatial wavenumber and ω is the temporal frequency. Let $\theta(\cdot)$ be an odd function, and let ϕ be a unit vector in \mathbb{R}^d . Let also $g(\lambda, \omega)$ be strictly positive, and let $g(\lambda, \omega)$ satisfy (3). If $g(\lambda - \theta(\omega)\phi, \omega)$ satisfies (3), then $\theta(\omega)$ must be locally bounded.

Proof We prove this statement using contradiction. Assume $\theta(\omega)$ is not locally bounded. The statement then follows almost directly from Lemma 2. To meet the conditions of Lemma 2, we need to define a triplet, $(\eta(\cdot), \xi_0(\cdot), U_0(\cdot))$ that has the properties listed in Lemma 2.

First we define $\xi_0(x)$. Since $\theta(\omega)$ is not locally bounded, there exists a sequence, $\omega_1, \omega_2, \dots$, and a bound, A , such that $|\omega_i| < A$ for all i and $\|\theta(\omega)\| \rightarrow \infty$ as $i \rightarrow \infty$. Define $\omega_0(x) = \omega_i$ if $i - 1 < x \leq i$ and $\omega_0 = 0$ otherwise. We have $|\omega_0(x)| < A$ for all $x \in \mathbb{R}$, and we

have that $\lim_{x \rightarrow \infty} \|\theta(\omega_0(x))\| \rightarrow \infty$. Define $\xi_0(x) = (\theta(\omega_0(x))\phi, \omega_0(x))$. Thus $\|\xi_0(x)\| \rightarrow \infty$ as $x \rightarrow \infty$.

Next, define $\eta(\lambda, \omega) = (\lambda - \theta(\omega)\phi, \omega)$. Consider $\eta(\xi_0(x)) = (\theta(\omega_0(x))\phi - \theta(\omega_0(x))\phi, \omega_0(x)) = (0, \omega_0(x))$. We know $|\omega_0(x)| < A$; therefore, $\|\eta(\xi_0(x))\|$ is bounded.

Finally we define $U_0(y) = (y\phi, 0)$. We prove the triplet has property (iii) in Lemma 2. Let $C > 0$. Consider $\eta(\xi_0(x) + U_0(y)) = (U_0(y), \omega_0(x)) = (y\phi, \omega_0(x))$. Therefore, $\|\eta(\xi_0(x) + U_0(y))\| \geq \|y\phi\| = y$. Let $Y = C$. If $y > Y$, it follows that $\|\eta(\xi_0(x) + U_0(y))\| > C$ for all $x > 0$.

By assumption, $g(\xi) = g(\lambda, \omega)$ satisfies (3), and above, we established the triplet, $(\eta(\cdot), \xi_0(\cdot), U_0(\cdot))$, has the properties in Lemma 2. Therefore, by Lemma 2, we know $g(\eta(\xi)) = g(\lambda - \theta(\omega)\phi, \omega)$ cannot satisfy (3). By contradiction, $\theta(\omega)$ must be locally bounded.

6.3 Half-Spectrum for Example 3

Writing down the half-spectrum in Example 3 requires expressions for the quantile function (inverse distribution function) of $|A| \sim 1/\sqrt{2\chi_{2\nu}^2}$ and the distribution function of the density $f_B \propto (\beta^2 + \omega^2)^{-(\kappa+1/2)}$. Write $P(s, x) = \gamma(s, x)/\Gamma(s)$ as the regularized gamma function, where γ is the lower-incomplete gamma function (Olver et al. 2010). The distribution function $F_{\chi_{2\nu}^2}(x; \nu)$ is given by $P(\nu, x/2)$. Hence $F_{\chi_{2\nu}^2}^{-1}(x; \nu) = 2P^{-1}(\nu, x)$, and therefore, $F_{|A|}^{-1} = 1/\sqrt{2F_{\chi_{2\nu}^2}^{-1}(x; \nu)}$. The inverse distribution function F_A^{-1} can be found directly as

$$F_A^{-1}(x; \nu) = \begin{cases} -1/\sqrt{4P^{-1}(\nu, 1-2x)}; & x < 1/2 \\ 1/\sqrt{4P^{-1}(\nu, 2x-1)}; & x \geq 1/2 \end{cases}.$$

To find F_B it is convenient to view f_B as a special case of the hypergeometric function, ${}_2F_1$ (Olver et al. 2010). We write $f_B \propto \beta^{-(2\kappa+1)} {}_2F_1(1, \kappa+1/2; 1; -\omega^2/\beta^2)$ (Olver et al. 2010). For $|\omega/\beta| < 1$, indefinite integration can be carried out directly using the series representation of the hypergeometric function to find $\int f_B d\omega \propto \omega/\beta^{2\kappa+1} {}_2F_1(1/2, \kappa+1/2; 3/2; -\omega^2/\beta^2)$ plus some constant. Hence by analytic continuation,

$$F_B(\omega; \beta, \kappa) = \frac{1}{2} + \frac{\Gamma(\kappa+1/2)}{\Gamma(\kappa)\sqrt{\pi}} \frac{\omega}{\beta} {}_2F_1\left(\frac{1}{2}, \frac{1}{2} + \kappa; \frac{3}{2}; -\frac{\omega^2}{\beta^2}\right)$$

can be defined for the values of ω outside this range (Olver et al. 2010). The constant in f_B and F_B can be obtained in (Gradshteyn & Ryzhik 2007, 3.241.4).

Define $\delta(\omega; \beta; \nu; \kappa) = |F_A^{-1}(F_B(\omega; \beta, \kappa); \nu)|$. With $\mathbb{C}(s; \alpha) = \phi \exp(-\alpha^2 s^2)$, we find the half spectral form of Example 3,

$$f(\omega)\mathbb{C}(s\delta(\omega)) = \phi(\beta^2 + \omega^2)^{-(\kappa+1/2)} \exp(-\alpha^2 s^2 \delta^2(\omega; \beta; \nu; \kappa)).$$

6.4 Half-Spectrum of K_{f_G}

To find the half-spectrum of K_{f_G} , we need only find the spectrum of the temporal covariance function $G(0, t)|_{\kappa=1} = \phi(\beta|u| + 1)^{-1}$ written without the spatial nugget in (11) and

equivalently in (10). (Bateman 1954, p. 8) gives the calculation as a cosine transform,

$$f_G(\omega) = \phi/\beta \left[\pi \sin\left(\frac{|\omega|}{\beta}\right) - 2 \operatorname{Si}\left(\frac{|\omega|}{\beta}\right) \sin\left(\frac{|\omega|}{\beta}\right) - 2 \operatorname{Ci}\left(\frac{|\omega|}{\beta}\right) \cos\left(\frac{|\omega|}{\beta}\right) \right],$$

where Si and Ci are the Sine and Cosine integrals defined in Olver et al. (2010). The tail of this spectrum decays asymptotically proportional to ω^{-2} (Olver et al. 2010). With f_G , the half-spectrum of K_{f_G} is obtained by plugging f_G into (6).

6.5 Fitting Procedures

R is used for fitting of all models using the general optimizer `nlm` with default optimization parameters. Numerical hessians are calculated in this implementation of `nlm`. In the previous two sections of this Appendix, we showed how half-spectra can be calculated provided functions ${}_2F_1$, Sine and Cosine integrals and the regularized gamma function, P , can be calculated. For P , we used the $\chi^2_{2\nu}$ quantile function included with *R*. We use the `gsl` package for the Sine and Cosine integral functions (Hankin 2006), and we used the `hypergeo` package for ${}_2F_1$ (Hankin & Lee 2006). The Block-Toeplitz solver for space-time fitting was adapted from a MATLAB function by Pepper & Fricke (2011).

References

- Akaike, H. (1973), ‘Block toeplitz matrix inversion’, *SIAM J. Appl. Math.* **24**, 234–241.
- Bateman, H. (1954), *Tables of Integral Transforms*, Vol. I, McGraw Hill.
- Brown, P., Kåresen, K., Roberts, G. & Tonellato, S. (2000), ‘Blur-generated non-separable space-time models’, *J. R. Statist. Soc. Ser. B* **62**, .
- Chilès, J. & Delfiner, P. (1999), *Geostatistics: Modeling Spatial Uncertainty*, Vol. , Wiley.
- Cressie, N. & Huang, H.-C. (1999), ‘Classes of nonseparable, spatio-temporal stationary covariance functions’, *J. Amer. Statist. Assoc.* , 1330–1340.
- de Luna, X. & Genton, M. (2005), ‘Predictive spatio-temporal models for spatially sparse environmental data’, *Statistica Sinica* **15**, 547–568.
- Fuentes, M., Chen, L. & Davis, J. (2007), ‘A class of nonseparable and nonstationary spatial temporal covariance functions’, *Environmetrics* **19**, 487–507.
- Genton, M. (2007), ‘Separable approximations of space-time covariance matrices’, *Environmetrics* **18**, 681–695.
- Gneiting, T. (2002), ‘Nonseparable, stationary covariance functions for spacetime data’, *J. Amer. Statist. Assoc.* **97**, .
- Gradshteyn, I. & Ryzhik, I. M. (2007), *Table of Integrals, Series, and Products*, Vol. , Academic Press.

- Hankin, R. K. S. (2006), ‘Special functions in r: introducing the gsl package’, *R News* **6**, .
- Hankin, R. K. S. & Lee, A. (2006), ‘A new family of non-negative distributions’, *Aust. and New Zeal. J. Stat.* **48**, 67–78.
- Haslett, J. & Raftery, A. (1989), ‘Space-time modeling with long-memory dependence: Assessing ireland’s wind power resource’, *Appl. Stat.* **38**, 1–50.
- Jun, M. & Stein, M. L. (2007), ‘An approach to producing space-time covariance functions on spheres’, *Technometrics* **49**, 468–479.
- Kent, J., Mohammadzadeh, M. & Mosammam, A. (2011), ‘The dimple in gneiting’s spatial-temporal covariance model’, *Biometrika* **98**, 489–494.
- Kent, J. & Mosammam, A. (Submitted), ‘Estimation and testing for covariance-spectral spatial-temporal models’, *arXiv.org*, *1409.4512* .
- Kolovos, A., Christakos, G., Hristopulos, D. & Serre, M. (2004), ‘Methods for generating non-separable spatiotemporal covariance models with potential environmental applications’, *Advances in Water Resources* **27**, 815–830.
- Kyriakidis, P. & Journel, A. (1999), ‘Geostatistical space-time models: a review’, *Mathematical Geology* **31**, 651–684.
- Lindgren, F., Rue, H. & Lindström, J. (2011), ‘An explicit link between gaussian fields and gaussian markov random fields: the spde approach’, *J. R. Statist. Soc. B* **73**, 423–498.
- Lindsay, B. (1988), ‘Composite likelihood methods’, *Contemporary Mathematics* **80**, 220–239.
- Ma, C. (2003), ‘Families of spatio-temporal stationary covariance models’, *J. Stat. Plan. and Infer.* **116**, 489–501.
- Olver, F., Lozier, D., Boisvert, R. & Clark, C. (2010), *NIST Handbook of Mathematical Functions*, Vol. , National Institute of Standards and Technology and Cambridge University Press.
- Pepper, K. & Fricke, T. (2011), ‘Block levinson solver: File id #30931’, *MATLAB Central* , .
- Priestley, M. (1983), *Spectral analysis and time series. Volumes 1 and 2*, Vol. , Academic Press.
- Sampson, P. & Guttorp, P. (1992), ‘Nonparametric estimation of nonstationary spatial covariance structure’, *J. Amer. Statist. Assoc.* **87**, 108–119.
- Schoenberg, I. (1938), ‘Metric spaces and completely monotone functions’, *Ann. Math.* **39**, 811–841.
- Stein, E. & Weiss, G. (1971), *Introduction to Fourier Analysis on Euclidean Spaces*, Vol. , Princeton University Press.
- Stein, M. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, Vol. , Springer.
- Stein, M. (2002), ‘The screening effect in kriging’, *The Annals of Statistics* **30**, 298–323.
- Stein, M. (2005a), ‘Space-time covariance functions’, *J. Amer. Statist. Assoc.* **100**, 310–321.

- Stein, M. (2005*b*), ‘Statistical methods for regular monitoring data’, *J. R. Statist. Soc. Ser. B* **67**, 667–687.
- Stein, M. (2011), ‘2010 rietz lecture: When does the screening effect hold?’, *Ann. Stat.* **39**, 2795–2819.
- Stein, M. (2015), ‘When does the screening effect not hold?’, *Spatial Statistics* **11**, 65–80.
- Stein, M., Chi, Z. & Welty, L. (2004), ‘Approximating likelihoods for large spatial data sets’, *J. R. Statist. Soc. B* **66**, 275–296.
- Vecchia, A. V. (1988), ‘Estimation and model identification for continuous spatial processes’, *J. R. Statist. Soc. B* **50**, 297–312.